

## **RAW SEQUENCE LISTING**

**The Biotechnology Systems Branch of the Scientific and Technical  
Information Center (STIC) no errors detected.**

Application Serial Number: 10/524,505  
Source: IFWP  
Date Processed by STIC: 6/26/06

# ***ENTERED***



IFWP

## RAW SEQUENCE LISTING

DATE: 06/26/2006

PATENT APPLICATION: US/10/524,505

TIME: 13:34:06

Input Set : A:\seq list.txt

Output Set: N:\CRF4\06262006\J524505.raw

1 <110> APPLICANT: NATIONAL INSTITUTE OF ADVANCED INDUSTRIAL SCIENCE AND TECHNOLOGY  
 2 AMERSHAM BIOSCIENCES KK  
 4 <120> TITLE OF INVENTION: Novel acetylgalactosamine transferases and nucleic acids  
 encoding  
 5 the same  
 7 <130> FILE REFERENCE: 159-86  
 9 <140> CURRENT APPLICATION NUMBER: 10/524,505  
 10 <141> CURRENT FILING DATE: 2005-02-14  
 13 <150> PRIOR APPLICATION NUMBER: PCT/JP03/10309  
 14 <151> PRIOR FILING DATE: 2003-08-13  
 17 <150> PRIOR APPLICATION NUMBER: JP2002-236292  
 18 <151> PRIOR FILING DATE: 2002-08-14  
 20 <160> NUMBER OF SEQ ID NOS: 34  
 22 <210> SEQ ID NO: 1  
 23 <211> LENGTH: 1039  
 24 <212> TYPE: PRT  
 25 <213> ORGANISM: Homo sapiens  
 27 <400> SEQUENCE: 1  
 28 Met Pro Arg Leu Pro Val Lys Lys Ile Arg Lys Gln Met Lys Leu Leu  
 29 1 5 10 15  
 31 Leu Leu Leu Leu Leu Ser Cys Ala Ala Trp Leu Thr Tyr Val His  
 32 20 25 30  
 34 Leu Gly Leu Val Arg Gln Gly Arg Ala Leu Arg Gln Arg Leu Gly Tyr  
 35 35 40 45  
 37 Gly Arg Asp Gly Glu Lys Leu Thr Ser Glu Thr Asp Gly Arg Gly Val  
 38 50 55 60  
 40 His Ala Ala Pro Ser Thr Gln Arg Ala Glu Asp Ser Ser Glu Ser Arg  
 41 65 70 75 80  
 43 Glu Glu Glu Gln Ala Pro Glu Gly Arg Asp Leu Asp Met Leu Phe Pro  
 44 85 90 95  
 46 Gly Gly Ala Gly Arg Leu Pro Leu Asn Phe Thr His Gln Thr Pro Pro  
 47 100 105 110  
 49 Trp Arg Glu Glu Tyr Lys Gly Gln Val Asn Leu His Val Phe Glu Asp  
 50 115 120 125  
 52 Trp Cys Gly Gly Ala Val Gly His Leu Arg Arg Asn Leu His Phe Pro  
 53 130 135 140  
 55 Leu Phe Pro His Thr Arg Thr Thr Val Lys Lys Leu Ala Val Ser Pro  
 56 145 150 155 160  
 58 Lys Trp Lys Asn Tyr Gly Leu Arg Ile Phe Gly Phe Ile His Pro Ala  
 59 165 170 175  
 61 Arg Asp Gly Asp Val Gln Phe Ser Val Ala Ser Asp Asp Asn Ser Glu  
 62 180 185 190  
 64 Phe Trp Leu Ser Leu Asp Glu Ser Pro Ala Ala Ala Gln Leu Val Ala  
 65 195 200 205

## RAW SEQUENCE LISTING

DATE: 06/26/2006

PATENT APPLICATION: US/10/524,505

TIME: 13:34:06

Input Set : A:\seq list.txt

Output Set: N:\CRF4\06262006\J524505.raw

```

67 Phe Val Gly Lys Thr Gly Ser Glu Trp Thr Ala Pro Gly Glu Phe Thr
68      210                      215                      220
70 Lys Phe Ser Ser Gln Val Ser Lys Pro Arg Arg Leu Met Ala Ser Arg
71 225                      230                      235                      240
73 Arg Tyr Tyr Phe Glu Leu Leu His Lys Gln Asp Asp Arg Gly Ser Asp
74      245                      250                      255
76 His Val Glu Val Gly Trp Arg Ala Phe Leu Pro Gly Leu Lys Phe Glu
77      260                      265                      270
79 Val Ile Ser Ser Ala His Ile Ser Leu Tyr Thr Asp Glu Ser Ala Leu
80      275                      280                      285
82 Lys Met Asp His Val Ala His Val Pro Gln Ser Pro Ala Ser His Val
83      290                      295                      300
85 Gly Gly Arg Pro Pro Gln Glu Glu Thr Ser Ala Asp Met Leu Arg Pro
86 305                      310                      315                      320
88 Asp Pro Arg Asp Thr Phe Phe Leu Thr Pro Arg Met Glu Ser Ser Ser
89      325                      330                      335
91 Leu Glu Asn Val Leu Glu Pro Cys Ala Tyr Ala Pro Thr Tyr Val Val
92      340                      345                      350
94 Lys Asp Phe Pro Ile Ala Arg Tyr Gln Gly Leu Gln Phe Val Tyr Leu
95      355                      360                      365
97 Ser Phe Val Tyr Pro Asn Asp Tyr Thr Arg Leu Thr His Met Glu Thr
98      370                      375                      380
100 Asp Asn Lys Cys Phe Tyr Arg Glu Ser Pro Leu Tyr Leu Glu Arg Phe
101 385                      390                      395                      400
103 Gly Phe Tyr Lys Tyr Met Lys Met Asp Lys Glu Glu Gly Asp Glu Asp
104      405                      410                      415
106 Glu Glu Asp Glu Val Gln Arg Arg Ala Phe Leu Phe Leu Asn Pro Asp
107      420                      425                      430
109 Asp Phe Leu Asp Asp Glu Asp Glu Gly Glu Leu Leu Asp Ser Leu Glu
110      435                      440                      445
112 Pro Thr Glu Ala Ala Pro Pro Arg Ser Gly Pro Gln Ser Pro Ala Pro
113      450                      455                      460
115 Ala Ala Pro Ala Gln Pro Gly Ala Thr Leu Ala Pro Pro Thr Pro Pro
116 465                      470                      475                      480
118 Arg Pro Arg Asp Gly Gly Thr Pro Arg His Ser Arg Ala Leu Ser Trp
119      485                      490                      495
121 Ala Ala Arg Ala Ala Arg Pro Leu Pro Leu Phe Leu Gly Arg Ala Pro
122      500                      505                      510
124 Pro Pro Arg Pro Ala Val Glu Gln Pro Pro Pro Lys Val Tyr Val Thr
125      515                      520                      525
127 Arg Val Arg Pro Gly Gln Arg Ala Ser Pro Arg Ala Pro Ala Pro Arg
128      530                      535                      540
130 Ala Pro Trp Pro Pro Phe Pro Gly Val Phe Leu His Pro Arg Pro Leu
131 545                      550                      555                      560
133 Pro Arg Val Gln Leu Arg Ala Pro Pro Arg Pro Pro Arg Pro His Gly
134      565                      570                      575
136 Arg Arg Thr Gly Gly Pro Gln Ala Thr Gln Pro Arg Pro Pro Ala Arg
137      580                      585                      590
139 Ala Gln Ala Thr Gln Gly Gly Arg Glu Gly Gln Ala Arg Thr Leu Gly

```

## RAW SEQUENCE LISTING

DATE: 06/26/2006

PATENT APPLICATION: US/10/524,505

TIME: 13:34:06

Input Set : A:\seq list.txt

Output Set: N:\CRF4\06262006\J524505.raw

```

140          595          600          605
142 Pro Ala Ala Pro Thr Val Asp Ser Asn Leu Ser Ser Glu Ala Arg Pro
143          610          615          620
145 Val Thr Ser Phe Leu Ser Leu Ser Gln Val Ser Gly Pro Gln Leu Pro
146 625          630          635          640
148 Gly Glu Gly Glu Glu Glu Glu Glu Gly Glu Asp Asp Gly Ala Pro Gly
149          645          650          655
151 Asp Glu Ala Ala Ser Glu Asp Ser Glu Glu Ala Ala Gly Pro Ala Leu
152          660          665          670
154 Gly Arg Trp Arg Glu Asp Ala Ile Asp Trp Gln Arg Thr Phe Ser Val
155          675          680          685
157 Gly Ala Val Asp Phe Glu Leu Leu Arg Ser Asp Trp Asn Asp Leu Arg
158          690          695          700
160 Cys Asn Val Ser Gly Asn Leu Gln Leu Pro Glu Ala Glu Ala Val Asp
161 705          710          715          720
163 Val Thr Ala Gln Tyr Met Glu Arg Leu Asn Ala Arg His Gly Gly Arg
164          725          730          735
166 Phe Ala Leu Leu Arg Ile Val Asn Val Glu Lys Arg Arg Asp Ser Ala
167          740          745          750
169 Arg Gly Ser Arg Phe Leu Leu Glu Leu Glu Leu Gln Glu Arg Gly Gly
170          755          760          765
172 Gly Arg Leu Arg Leu Ser Glu Tyr Val Phe Leu Arg Leu Pro Gly Ala
173          770          775          780
175 Arg Val Gly Asp Ala Asp Gly Glu Ser Pro Glu Pro Ala Pro Ala Ala
176 785          790          795          800
178 Ser Val Arg Pro Asp Gly Arg Pro Glu Leu Cys Arg Pro Leu Arg Leu
179          805          810          815
181 Ala Trp Arg Gln Asp Val Met Val His Phe Ile Val Pro Val Lys Asn
182          820          825          830
184 Gln Ala Arg Trp Val Ala Gln Phe Leu Ala Asp Met Ala Ala Leu His
185          835          840          845
187 Ala Arg Thr Gly Asp Ser Arg Phe Ser Val Val Leu Val Asp Phe Glu
188          850          855          860
190 Ser Glu Asp Met Asp Val Glu Arg Ala Leu Arg Ala Ala Arg Leu Pro
191 865          870          875          880
193 Arg Tyr Gln Tyr Leu Arg Arg Thr Gly Asn Phe Glu Arg Ser Ala Gly
194          885          890          895
196 Leu Gln Ala Gly Val Asp Ala Val Glu Asp Ala Ser Ser Ile Val Phe
197          900          905          910
199 Leu Cys Asp Leu His Ile His Phe Pro Pro Asn Ile Leu Asp Gly Ile
200          915          920          925
202 Arg Lys His Cys Val Glu Gly Arg Leu Ala Phe Ala Pro Val Val Met
203          930          935          940
205 Arg Leu Ser Cys Gly Ser Ser Pro Arg Asp Pro His Gly Tyr Trp Glu
206 945          950          955          960
208 Val Asn Gly Phe Gly Leu Phe Gly Ile Tyr Lys Ser Asp Phe Asp Arg
209          965          970          975
211 Val Gly Gly Met Asn Thr Glu Glu Phe Arg Asp Gln Trp Gly Gly Glu
212          980          985          990

```

## RAW SEQUENCE LISTING

DATE: 06/26/2006

PATENT APPLICATION: US/10/524,505

TIME: 13:34:06

Input Set : A:\seq list.txt

Output Set: N:\CRF4\06262006\J524505.raw

```

214 Asp Trp Glu Leu Leu Asp Arg Val Leu Gln Ala Gly Leu Glu Val Glu
215      995      1000      1005
217 Arg Leu Arg Leu Arg Asn Phe Tyr His His Tyr His Ser Lys Arg Gly
218      1010      1015      1020
220 Met Trp Ser Val Arg Ser Arg Lys Gly Ser Arg Thr Gly Ala Ser
221 1025      1030      1035      1039
224 <210> SEQ ID NO: 2
225 <211> LENGTH: 3120
226 <212> TYPE: DNA
227 <213> ORGANISM: Homo sapiens
229 <400> SEQUENCE: 2
230 atg ccg cgg ctg ccg gtg aag aag atc cgt aag cag atg aag ctg ctg      48
231 Met Pro Arg Leu Pro Val Lys Lys Ile Arg Lys Gln Met Lys Leu Leu
232 1      5      10      15
234 ctg ctg ctg ctg ctg ctg agc tgc gcc gcg tgg ctg acc tac gtg cac      96
235 Leu Leu Leu Leu Leu Leu Ser Cys Ala Ala Trp Leu Thr Tyr Val His
236      20      25      30
238 ctg ggc ctg gtg cgc cag gga cgc gcg ctg cgc cag cgc ctg ggc tac      144
239 Leu Gly Leu Val Arg Gly Gly Arg Ala Leu Arg Gln Arg Leu Gly Tyr
240      35      40      45
242 ggg cga gat ggt gag aag ctg acc agt gag acc gac ggc cgg ggg gtc      192
243 Gly Arg Asp Gly Glu Lys Leu Thr Ser Glu Thr Asp Gly Arg Gly Val
244      50      55      60
246 cac gct gcg cca tcc aca cag agg gct gag gac tcc agt gag agc cgt      240
247 His Ala Ala Pro Ser Thr Gln Arg Ala Glu Asp Ser Ser Glu Ser Arg
248 65      70      75      80
250 gaa gag gag caa gcg ccc gaa ggt cgg gac cta gac atg ctg ttt cct      288
251 Glu Glu Glu Gln Ala Pro Glu Gly Arg Asp Leu Asp Met Leu Phe Pro
252      85      90      95
254 ggg ggg gct ggg agg ctg cca ctg aac ttc acc cat cag aca ccc cca      336
255 Gly Gly Ala Gly Arg Leu Pro Leu Asn Phe Thr His Gln Thr Pro Pro
256      100      105      110
258 tgg cgg gag gag tac aag ggg cag gtg aac ctg cac gtg ttt gag gac      384
259 Trp Arg Glu Glu Tyr Lys Gly Gln Val Asn Leu His Val Phe Glu Asp
260      115      120      125
262 tgg tgt ggg ggc gcc gtg ggc cac ctg agg agg aac ctg cac ttc ccg      432
263 Trp Cys Gly Gly Ala Val Gly His Leu Arg Arg Asn Leu His Phe Pro
264      130      135      140
266 ctg ttc cct cat acg cgc acc acc gtg aag aag ttg gcc gtg tcc ccc      480
267 Leu Phe Pro His Thr Arg Thr Thr Val Lys Lys Leu Ala Val Ser Pro
268 145      150      155      160
270 aag tgg aag aac tat gga ctg cgt att ttt ggt ttc atc cac ccg gcg      528
271 Lys Trp Lys Asn Tyr Gly Leu Arg Ile Phe Gly Phe Ile His Pro Ala
272      165      170      175
274 agg gac gga gac gtc cag ttt tct gtg gcc tca gac gac aac tcg gag      576
275 Arg Asp Gly Asp Val Gln Phe Ser Val Ala Ser Asp Asp Asn Ser Glu
276      180      185      190
278 ttc tgg ctg agt ctg gac gag agc cct gct gct gcc cag ctt gtg gcc      624
279 Phe Trp Leu Ser Leu Asp Glu Ser Pro Ala Ala Ala Gln Leu Val Ala

```

## RAW SEQUENCE LISTING

DATE: 06/26/2006

PATENT APPLICATION: US/10/524,505

TIME: 13:34:06

Input Set : A:\seq list.txt

Output Set: N:\CRF4\06262006\J524505.raw

280	195	200	205	
282	ttt gtg ggc aag act ggc tcc gag tgg aca gcg cct gga gaa ttc acc	672		
283	Phe Val Gly Lys Thr Gly Ser Glu Trp Thr Ala Pro Gly Glu Phe Thr			
284	210 215 220			
286	aag ttc agc tcc cag gtg tcc aag ccc agg cgg ctc atg gcc tcc cgg	720		
287	Lys Phe Ser Ser Gln Val Ser Lys Pro Arg Arg Leu Met Ala Ser Arg			
288	225 230 235 240			
290	agg tac tac ttt gag ttg ctg cac aag cag gac gac cgc ggc tcg gac	768		
291	Arg Tyr Tyr Phe Glu Leu Leu His Lys Gln Asp Asp Arg Gly Ser Asp			
292	245 250 255			
294	cac gtg gaa gtg ggc tgg cga gct ttc ctg ccc ggc ctg aag ttc gag	816		
295	His Val Glu Val Gly Trp Arg Ala Phe Leu Pro Gly Leu Lys Phe Glu			
296	260 265 270			
298	gtc atc agc tct gct cac atc tcc ctg tac aca gat gag tca gcc ttg	864		
299	Val Ile Ser Ser Ala His Ile Ser Leu Tyr Thr Asp Glu Ser Ala Leu			
300	275 280 285			
302	aag atg gac cac gtg gcg cac gtc ccc cag tct cca gcc agc cac gtg	912		
303	Lys Met Asp His Val Ala His Val Pro Gln Ser Pro Ala Ser His Val			
304	290 295 300			
306	ggg ggg cgt ccg ccg cag gag gag acc agc gca gac atg ctg cgg cca	960		
307	Gly Gly Arg Pro Pro Gln Glu Glu Thr Ser Ala Asp Met Leu Arg Pro			
308	305 310 315 320			
310	gat ccc agg gat acc ttt ttc ctc act cca cgc atg gaa tct tcg agc	1008		
311	Asp Pro Arg Asp Thr Phe Phe Leu Thr Pro Arg Met Glu Ser Ser Ser			
312	325 330 335			
314	ctg gag aac gtg ctg gag ccc tgc gcc tac gcc ccc acc tac gtg gtc	1056		
315	Leu Glu Asn Val Leu Glu Pro Cys Ala Tyr Ala Pro Thr Tyr Val Val			
316	340 345 350			
318	aag gac ttc ccg atc gcc aga tac cag ggc ctg caa ttt gtg tac ctg	1104		
319	Lys Asp Phe Pro Ile Ala Arg Tyr Gln Gly Leu Gln Phe Val Tyr Leu			
320	355 360 365			
322	tcc ttc gtt tat ccc aac gac tac act cgc ctc acc cac atg gag acg	1152		
323	Ser Phe Val Tyr Pro Asn Asp Tyr Thr Arg Leu Thr His Met Glu Thr			
324	370 375 380			
326	gac aac aag tgc ttc tac cgc gag tct ccg ctg tat ctg gag agg ttt	1200		
327	Asp Asn Lys Cys Phe Tyr Arg Glu Ser Pro Leu Tyr Leu Glu Arg Phe			
328	385 390 395 400			
330	ggg ttc tat aaa tac atg aag atg gac aag gag gag ggg gat gag gat	1248		
331	Gly Phe Tyr Lys Tyr Met Lys Met Asp Lys Glu Glu Gly Asp Glu Asp			
332	405 410 415			
334	gaa gaa gac gag gtg cag cgc cga gcc ttc ctc ttc ctc aac ccg gac	1296		
335	Glu Glu Asp Glu Val Gln Arg Arg Ala Phe Leu Phe Leu Asn Pro Asp			
336	420 425 430			
338	gac ttc ctg gac gac gag gac gag ggg gag ctg ctc gac agc ctg gag	1344		
339	Asp Phe Leu Asp Asp Glu Asp Glu Gly Glu Leu Leu Asp Ser Leu Glu			
340	435 440 445			
342	ccc acc gag gcg gcc ccg ccc agg agc ggc ccc cag tcc ccc gcc cca	1392		
343	Pro Thr Glu Ala Ala Pro Pro Arg Ser Gly Pro Gln Ser Pro Ala Pro			
344	450 455 460			

**VERIFICATION SUMMARY**

PATENT APPLICATION: US/10/524,505

DATE: 06/26/2006

TIME: 13:34:07

Input Set : A:\seq list.txt

Output Set: N:\CRF4\06262006\J524505.raw